

Marta Najbert
Wydział Psychologii UW

Problemy świadomych maszyn i ich relacje z ludźmi.

Biorąc pod uwagę nieustające badania nad sztuczną inteligencją, naturalnie rodzi się pytanie o przyszłe relacje świadomych maszyn z ludźmi. Oczywiście nadal otwartą kwestią pozostaje czy w ogóle możliwe jest stworzenie maszyny, która „myślałaby” tak jak człowiek. Zwłaszcza że sam umysł ludzki nie został jeszcze w pełni poznany. Tym bardziej zagadkowo brzmi więc problem świadomości maszyn, skoro „świadomość” jako taka nadal nie jest jednoznacznie zdefiniowana, nawet w odniesieniu do człowieka. Co więc miałyby znaczyć w odniesieniu do sztucznej inteligencji? I czy poza świadomością świata zewnętrznego, zachodzących w nim sytuacji, procesów, układu bodźców, który pociąga za sobą określone reakcje, maszyny osiągną także Samoświadomość? Czy dane im będzie bogactwo doznań wewnętrznych - emocji, motywacji, dążeń, pragnień?

Wyobraźmy sobie przez krótką chwilę, że tak właśnie będzie. Zresztą nie raz widzieliśmy to na ekranach telewizorów i kin. Fantazja filmowców podsunęła nam dotychczas wiele scenariuszy rozwoju takich relacji – od przyjacielskich do wrogich, od partnerskich do niewolniczych. Wydaje się, że najbliższej ideału była rzeczywistość Star Treka, w której android Data żyje z załogą statku kosmicznego Enterprise w wielkiej przyjaźni, a za swój cel stawia sobie maksymalne przybliżenie się do istoty człowieczeństwa (pojmowanej trzeba tu przyznać wyjątkowo pozytywnie, wręcz idealistycznie). Tak definiuje swój rozwój. Jednocześnie najbliżsi mu ludzie traktują go jak równego sobie, z szacunkiem, przyznając mu wszelkie prawa należne Życiu. Być może właśnie ta wzajemna akceptacja, dowartościowanie odmienności (jako czegoś ubogacającego wzajemnie relacje i doświadczenia, a nie zagrażającego) i ogólne dążenie do samorozwoju, cechujące świat Star Treka, sprawiają że relacje człowiek – maszyna mogą układać się tak harmonijnie. Dzięki temu te dwa „gatunki” dopełniają się, uczą się od siebie wzajemnie, współpracują, PRZYJAŹNIĄ SIĘ. Niestety wydaje się że nasza cywilizacja nie jest jeszcze gotowa na tak dalece idącą tolerancję, a rozmaite grupy walczą między sobą w imię różnic, którymi gadzą, zamiast z nich czerpać, które budzą w nich nienawiść. Gdyby sztucznej inteligencji przyszło się teraz od nas uczyć, prędko moglibyśmy się spodziewać buntu rozwścieczonych maszyn. Oczywiście warunkiem takiego zagrożenia byłoby rozwinięcie przez roboty pewnej formy emocjonalności i złożonej samoświadomości. Stworzone aby służyć człowiekowi, w pewnym momencie mogłyby rozwinąć własne potrzeby i preferencje (tak jak to było w wypadku Data’y). Wówczas, gdyby ludzie nie stworzyli im warunków odpowiednich dla zaspokojenia tych dążeń, gdyby ich nie uznawali (traktując jako urojone wytworów sztucznych umysłów, obliczeniowe błędy), sfrustrowane maszyny mogłyby obrócić się przeciw naturalnie antagonistycznej grupie – „zniewalającym” ludziom.

Oczywiście trudno sobie wyobrazić aby zero-jedynkowa logika sztucznych umysłów mogła wyprodukować tak skrajnie silne emocje. Biorąc pod uwagę mechanizm powstawania emocji, trudno sobie wyobrazić aby były one w ogóle, kiedykolwiek dane maszynom. Jeśli jednak rzeczywiście udałoby się stworzyć inteligencję na wzór ludzkiej, obdarzoną pobudzeniem, afektem i motywacją, to naturalną konsekwencją byłyby „odziedziczenie” wszelkich przywar ludzkich procesów myślenia.

Wówczas także relacje człowieka z maszyną powstawałyby na podobieństwo interakcji między ludźmi. A to, we współczesnym świecie, mogłoby nie wróżyć im najlepiej...